AMENDMENTS TO THE CLAIMS:

This listing of the claims will replace all prior versions, and listings, of the claims in this application.

Listing of Claims:

1. (Currently Amended) A method to process a document, comprising:

partitioning, with a tokenizer, document text separated by spaces into a plurality of tokens based on the spaces;

identifying tokens to be ignored and not considered;

determining that a first token considered of the plurality of tokens comprises a chemical name fragment, wherein determining comprises:

examining syntax of the first token,

examining context of the first token with respect to at least one adjacent token of the plurality of tokens, and

taking into account the syntax and the context, applying to the first token a plurality of regular expressions, rules, and a plurality of dictionaries comprised of a prefix dictionary, and a suffix dictionary to recognize the chemical name fragments;

adding the recognized chemical name fragment to a vector of chemical name fragments, where the chemical name fragment is identified by a vector index variable;

combining the ~~first token~~ recognized chemical name fragment with at least one of the adjacent tokens that are determined to be a chemical name fragment into a complete chemical name, where combining comprises:

initializing the chemical name fragment vector index variable,

incrementing the chemical name fragment vector index variable, where the incrementing continues at least until no chemical name fragments remain;

setting a string combination to include the chemical name fragments identified by the initialized and incremented chemical name fragment vector index variables, and

adding the string combination to a vector c as the complete chemical name;

assigning the complete chemical name with one part of speech; and

storing in a memory the complete chemical name assigned with the one part of speech~~;~~

~~where identifying tokens to be ignored comprises applying a negative dictionary to the plurality of tokens and wherein the plurality of dictionaries consists of the prefix dictionary, the suffix dictionary, and the negative dictionary~~.

2. (Original) A method as in claim 1, where the complete chemical name is assigned a noun phrase part of speech.

3-4. (Canceled)

5. (Original) A method as in claim 1, further comprising filtering recognized chemical name fragments using a list of stop words to eliminate erroneous chemical name fragments.

6. (Original) A method as in claim 1, where chemical name fragments are further recognized by using common chemical word endings.

7. (Original) A method as in claim 1, where application of said regular expressions and rules results in punctuation characters being one of maintained or removed between chemical name fragments as a function of context.

8. (Original) A method as in claim 1, where said regular expressions comprise a plurality of patterns, individual ones of which are comprised of at least one of characters, numbers and punctuation.

9. (Original) A method as in claim 8, where the punctuation comprises at least one of parenthesis, square bracket, hyphen, colon and semi-colon.

10. (Currently Amended) A method as in claim 8, where the characters comprise ~~at least one of~~ upper case C, O, R, N and H, and where the characters comprise strings of lower case xy, ene, ine, yl, ane and oic.

11. (Currently Amended) A method as in ~~claim 8~~ claim 1, ~~where the characters comprise strings of at least one of lower case xy, ene, ine, yl, ane and oic~~ where identifying tokens to be ignored comprises applying a negative dictionary to the plurality of tokens and wherein the plurality of dictionaries consists of the prefix dictionary, the suffix dictionary, and the negative dictionary.

12. (Original) A method as in claim 1, comprising an initial step of tokenizing the document to provide a sequence of tokens.

13. (Currently Amended) A system for processing a text document, comprising:

a first unit for partitioning document text separated by spaces into a plurality of tokens based on the spaces;

a second unit, operable for identifying tokens to be ignored and not considered;

a third unit, operable for determining that a first token considered of the plurality of tokens comprises a chemical name fragment, wherein determining comprises:

examining context of the first token with respect to at least one adjacent token of the plurality of tokens, and

taking into account the syntax and the context, applying to the first token a plurality of regular expressions, rules and a plurality of dictionaries comprised of a prefix dictionary and a suffix dictionary to recognize the chemical name fragment~~s~~;

a fourth unit operable, to add the recognized chemical name fragment to a vector of chemical name fragments, where the chemical name fragment is identified by a vector index variable;

a ~~fourth~~ fifth unit operable, to combine the ~~first token~~ recognized chemical name fragment with at least one of the adjacent tokens that are determined to be a chemical name fragment, wherein:

the fifth unit is operable to initialize the chemical name fragment vector index variable,
the fifth unit is operable to increment the chemical name fragment vector index variable, where the incrementing continues at least until no chemical name fragments remain;
the fifth unit is operable to set a string combination to include the chemical name fragments identified by the initialized and incremented chemical name fragment vector index variables, and
the fifth unit is operable to add the string combination to a vector c as the complete chemical name;

a ~~fifth~~ sixth unit operable to assign the complete chemical name with one part of speech; and

a ~~sixth~~ seventh unit operable for storing in a memory the complete chemical name assigned with one part of speech~~;~~

~~where the second unit is operable for identifying the tokens to be ignored by applying a negative dictionary to the plurality of tokens, and wherein the plurality of dictionaries consists of the prefix dictionary, the suffix dictionary, and the negative dictionary~~.

14. (Original) A system as in claim 13, where the complete chemical name is assigned a noun phrase part of speech.

15-16. (Canceled)

17. (Original) A system as in claim 13, where said second unit further comprises a sub-unit for filtering recognized chemical name fragments using a list of stop words to eliminate erroneous chemical name fragments.

18. (Original) A system as in claim 13, where chemical name fragments are further recognized by using common chemical word endings.

19. (Original) A system as in claim 13, where application of said regular expressions and rules results in punctuation characters being one of maintained or removed between chemical name fragments as a function of context.

20. (Original) A system as in claim 13, where said regular expressions comprise a plurality of patterns, individual ones of which are comprised of at least one of characters, numbers and punctuation.

21. (Original) A system as in claim 20, where the punctuation comprises at least one of parenthesis, square bracket, hyphen, colon and semi-colon.

22. (Currently Amended) A system as in claim 20, where the characters comprise ~~at least one of~~ upper case C, O, R, N and H, and where the characters comprise strings of lower case xy, ene, ine, yl, ane and oic.

23. (Currently Amended) A system as in ~~claim 20~~ claim 13, ~~where the characters comprise strings of at least one of lower case xy, ene, ine, yl, ane and oic~~ where the second unit is operable for identifying the tokens to be ignored by applying a negative dictionary to the plurality of tokens, and wherein the plurality of dictionaries consists of the prefix dictionary, the suffix dictionary, and the negative dictionary.

24. (Original) A system as in claim 13, further comprising a tokenizer for tokenizing the document to provide a sequence of tokens.

25. (Currently Amended) A computer program product ~~embodied on~~ comprising a memory that contains software ~~and~~ executable to perform operations, the operations comprising:

partitioning a document text separated by spaces into a plurality of tokens based on the spaces;

identifying tokens to be ignored and not considered;

determining that a first token considered of the plurality of tokens comprises ~~an organic~~ a chemical name fragment, wherein determining comprises:

examining syntax of the first token,

examining context of the first token with respect to at least one adjacent token of the plurality of tokens, and

taking into account the syntax and the context, applying a plurality of regular expressions, rules, and a plurality of dictionaries comprising a prefix dictionary, and a suffix dictionary to recognize ~~organic~~ the chemical name fragment~~s~~;

adding the recognized chemical name fragment to a vector of chemical name fragments, where the chemical name fragment is identified by a vector index variable;

combining the ~~first token~~ recognized chemical name fragment with at least one of the adjacent tokens that are determined to be ~~an organic~~ a chemical name fragment into a complete ~~organic~~ chemical name, where combining comprises:

initializing the chemical name fragment vector index variable,

incrementing the chemical name fragment vector index variable, where the incrementing continues at least until no chemical name fragments remain;

setting a string combination to include the chemical name fragments identified by the initialized and incremented chemical name fragment vector index variables, and

adding the string combination to a vector c as the complete chemical name;

assigning the complete ~~organic~~ chemical name with one part of speech; and

storing in a memory the complete ~~organic~~ chemical name with the one part of speech~~;~~

~~where identifying tokens to be ignored comprises applying a negative dictionary to the plurality of tokens and wherein the plurality of dictionaries consists of the prefix dictionary, the suffix dictionary, and the negative dictionary~~.

26. (Currently Amended) A computer program product as in claim 25, where the complete ~~organic~~ chemical name is assigned a noun phrase part of speech.

27-28. (Canceled)

29. (Currently Amended) A computer program product as in claim 25, further comprising instructions for filtering recognized ~~organic~~ chemical name fragments using a list of stop words to eliminate erroneous fragments.

30. (Original) A computer program product as in claim 25, where chemical name fragments are further recognized by using common chemical word endings.

31. (Currently Amended) A computer program product as in claim 25, where application of said regular expressions and rules results in punctuation characters being one of maintained or removed between ~~organic~~ chemical name fragments as a function of context.

32. (Original) A computer program product as in claim 25, where said regular expressions comprise a plurality of patterns, individual ones of which are comprised of at least one of characters, numbers and punctuation.

33. (Currently Amended) A computer program product as in claim 32, where the punctuation comprises at least one of parenthesis, square bracket, hyphen, colon and semi-colon, where

the characters comprise ~~at least one of~~ upper case C, O, R, N and H, and further comprise strings of ~~at least one of~~ lower case xy, ene, ine, yl, ane and oic.

34. (Currently Amended) A computer program product as in claim 25, where said instructions for assigning operate on a sequence of tokens derived from document text, and where identifying tokens to be ignored comprises applying a negative dictionary to the plurality of tokens and wherein the plurality of dictionaries consists of the prefix dictionary, the suffix dictionary, and the negative dictionary.

35. (Currently Amended) A system comprising a ~~plurality of computers at least two of which are coupled together through a data communications network~~ at least one computer, said system comprising

a first unit for partitioning document text separated by spaces into a plurality of tokens based on the spaces;

a second unit, operable for identifying tokens to be ignored and not considered;

a third unit, operable for determining that a first token considered of the plurality of tokens comprises a chemical name fragment, wherein determining comprises:

examining syntax of the first token,

examining context of the first token with respect to at least one adjacent token of the plurality of tokens, and

taking into account the syntax and the context, applying a plurality of regular expressions, rules, and a plurality of dictionaries comprised of a prefix dictionary and a syntax dictionary to recognize the chemical name fragment~~s~~;

a fourth unit, operable to add the recognized chemical name fragment to a vector of chemical name fragments, where the chemical name fragment is identified by a vector index variable;

a ~~fourth~~ fifth unit, operable to combine the ~~first token~~ recognized chemical name fragment with at least one of the adjacent tokens that are determined to be a chemical name fragment into a complete chemical name, where combining comprises:

the fifth unit is operable to initialize the chemical name fragment vector index variable,

the fifth unit is operable to increment the chemical name fragment vector index variable, where the incrementing continues at least until no chemical name fragments remain;

the fifth unit is operable to set a string combination to include the chemical name fragments identified by the initialized and incremented chemical name fragment vector index variables, and

the fifth unit is operable to add the string combination to a vector c as the complete chemical name;

a ~~fifth~~ sixth unit, operable to assign the complete chemical name with one part of speech; and

a ~~sixth~~ seventh unit, operable for storing in a memory information the complete chemical name with the one part of speech~~;~~

~~where the second unit is operable for identifying the tokens to be ignored by applying a negative dictionary to the plurality of tokens, and wherein the plurality of dictionaries consists of the prefix dictionary, the suffix dictionary, and the negative dictionary~~.

36. (Currently Amended) A system as in claim 35, where the complete chemical name is assigned a noun phrase part of speech, and where the second unit is operable for identifying the tokens to be ignored by applying a negative dictionary to the plurality of tokens, and wherein the plurality of dictionaries consists of the prefix dictionary, the suffix dictionary, and the negative dictionary.

37. (Original) A system as in claim 35, where a user of the system accesses the system through a data communications network.

38-39. (Canceled)